

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2003, 13:35:55 ; Search time 10 Seconds

(without alignments)
394.025 Million cell updates/sec

Title: US-09-806-302a-2

Sequence: 1 MKLLVLMIALLLHCHYADS.....NFGLMHIVYSIMCKMSKN 95

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	100.0	95	1	MG8B_HUMAN
2	12	12.6	93	1	MG8B_HUMAN
3	9	9.5	705	1	DPRO2_YEAST
4	8	8.4	552	1	NREPE_ECOLI
5	8	8.4	2105	1	POLR_ASGBP
6	7	7.4	112	1	UCN2_HUMAN
7	7	7.4	239	1	GIDB_CLOAB
8	7	7.4	377	1	TYRA_HAEIN
9	7	7.4	433	1	PROA_SYNY3
10	7	7.4	518	1	LEU1_BUCRP
11	7	7.4	943	1	ARS2_DROME
12	7	7.4	994	1	GLNE_MYCTU
13	6	6.3	79	1	Y8KD_STRLI
14	6	6.3	81	1	TFPI_RAT
15	6	6.3	95	1	PSC3_RAT
16	6	6.3	102	1	RL12_METJA
17	6	6.3	106	1	RL12_ARCFU
18	6	6.3	107	1	RL12_PRRAB
19	6	6.3	108	1	RL12_PRRHO
20	6	6.3	113	1	NLTA_WHEAT
21	6	6.3	124	1	UCN1_HUMAN
22	6	6.3	126	1	FMKH_MTCOL
23	6	6.3	132	1	Y125_AOUAE
24	6	6.3	154	1	PFDA_HALNI
25	6	6.3	154	1	SSRP_ENTFA
26	6	6.3	157	1	Y905_METTH
27	6	6.3	167	1	MUGH_BRALA
28	6	6.3	167	1	TYCN_HAEIN
29	6	6.3	175	1	ANP_OSMO
30	6	6.3	180	1	AROK_HAEIN
31	6	6.3	210	1	MTRR_NEIGO
32	6	6.3	211	1	YN03_CAEEL
33	6	6.3	215	1	MTCL_METMA

34	6	6.3	216	1	EPF1_XENLA
35	6	6.3	216	1	MTCL_METAC
36	6	6.3	216	1	MTTC_METBA
37	6	6.3	217	1	MTC2_METAC
38	6	6.3	218	1	MTC2_METMA
39	6	6.3	223	1	CPL_PEA
40	6	6.3	225	1	Y5A5_RHIME
41	6	6.3	235	1	IN21_MAIZE
42	6	6.3	243	1	PNR2_XYLLA
43	6	6.3	249	1	ETFE_BRALA
44	6	6.3	249	1	ETFE_BRALA
45	6	6.3	262	1	DRN1_PIG

ALIGNMENTS

RESULT 1
MG8B_HUMAN STANDARD: PRT: 95 AA.
AC 075556;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mammaglobin B precursor (Mammaglobin 2) (Lipophilin C) (Lactogloblin)
DE (Secretogloblin family 2A member 1).
GN SCG82A1 OR MG82 OR UCB3 OR LIPHC..
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=99167354; PubMed=10066439;
RA Zhao C., Nguyen T., Yustlov T., Glasgow B.J., Lohrer R.I.;
RT Becker R.M., Darlow C., Zimonjic D.B., Popescu H.C., Watson M.A.,
RA Fleming T.P.;
RT "Identification of mammaglobin B, a novel member of the uteroglobulin
RT gene family.";
RL Genomics 54:70-78(1998).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=99167354; PubMed=10066439;
RA Zhao C., Nguyen T., Yustlov T., Glasgow B.J., Lohrer R.I.;
RT "Lipophilins: human peptides homologous to rat prostatein.";
RL Biochem. Biophys. Res. Commun. 256:147-155(1999).
RN 13
RP SEQUENCE OF 19-85.
RC TISSUE=tears;
RX MEDLINE=9816342; PubMed=9504814;
RA Molloy M.P., Bolis S., Herbert B.R., Ou K., Tyler M.I., van Dyk D.D.,
RA Willcox M.D., Gooley A.A., Williams K.L., Morris C.A., Walsh B.J.;
RT "Establishment of the human reflex tear two-dimensional polycrylamide
RT gel electrophoresis reference map: new proteins of potential
RT diagnostic value.";
RL Electrophoresis 18:2811-2815(1997).
RN 14
RP SEQUENCE OF 19-46 AND 60-78, AND MASS SPECTROMETRY.
RX MEDLINE=98385871; PubMed=9720917;
RA Lehrer R.I., Xu G., Abduragimov A., Dinh N.N., Yu X.-D., Martin D.,
RA Glasgow B.J.;
RT "Lipophilin, a novel heterodimeric protein of human tears.";
RL FEBS Lett. 432:163-167(1998).
RP FUNCTION: MAY BIND ANDROGENS AND OTHER STEROIDS. MAY ALSO BIND
RP ESTRADIOL. A CHEMOTHERAPEUTIC AGENT USED FOR PROSTATE CANCER.
CC MAY BE UNDER TRANSCRIPTIONAL REGULATION OF STEROID HORMONES.
CC - SUBUNIT: HETERODIMER OF A LIPOPHILIN A AND A LIPOPHILIN C
CC (MAMMAGLOBIN B) MONOMER ASSOCIATED HEAD TO HEAD.
CC - TISSUE SPECIFICITY: EXPRESSED IN THYMUS, TRICHLA, KIDNEY, STEROID
CC RESPONSIVE TISSUES (PROSTATE, TESTIS, UTERUS, BREAST AND OVARY).
CC AND SALIVARY GLAND.
CC - MASS SPECTROMETRY: MW=8854.94; METHOD=Electrospray; RANGE=15-95.
CC - SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
CC SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF071219; AAC79996.1; -
 DR EMBL: AJ224173; CA11865.1; -
 DR Genew: HGNC:7051; SCGB2A1.
 DR MIM: 604398; -
 DR InterPro: IPR003627; Mamgb/prostatn.
 DR InterPro: IPR000329; Uteroglobln.
 DR Pfam: PF01099; Uteroglobln; 1.
 DR ProDom: PD029354; Mamgb/prostatn; 1.
 DR PROSITE: PS00403; UTEROGLOBIN_1; FALSE_NEG.
 DR PROSITE: PS00404; UTEROGLOBIN_2; FALSE_NEG.
 KW Signal; Glycoprotein.
 FT SIGNAL 1 18
 FT CHAIN 19 95 MAMAGLOBIN B.
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT SEQUENCE 95 AA; 10884 MW; 0719738289F89F8D CRC64;
 Query Match 100.0%; Score 95; DB 1; Length 95;
 Best Local Similarity 100.0%; Pred. No. 7.6e-90;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MKLLVLMALALLHCYADSGCKLLEDMVEKTSISFEYKELLOEFDSDAAACMG 60
 Db 1 MKLLVLMALALLHCYADSGCKLLEDMVEKTSISFEYKELLOEFDSDAAACMG 60
 Oy 61 KFKOCFLNOSHRTLKNGFLMHTVYDSIMCNKSN 95
 Db 61 KFKOCFLNOSHRTLKNGFLMHTVYDSIMCNKSN 95
 Db 61 KFKOCFLNOSHRTLKNGFLMHTVYDSIMCNKSN 95
 RESULT 2
 MGBA_HUMAN STANDARD; PRT; 93 AA.
 ID MGBA_HUMAN
 AC 013296;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Mammaglobin A precursor (Mammaglobin 1) (Secretoglobin family 2A member 2).
 GN SGB2A2 OR MGRI OR UGB2.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RX NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE-Breast;
 RA MEDLINE-96223598; PubMed-8631025;
 RA Watson M.A., Fleming T.P.;
 RT "Mammaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer.";
 RT Cancer Res. 56:860-865(1996).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98147371; PubMed-9486047;
 RA Watson M.A., Darlow C., Zimonjic D.B., Popescu N.C., Fleming T.;
 RT "Structure and transcriptional regulation of the human mammaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to chromosome 11q13.";
 RT Oncogene 16:817-824(1998).
 RL [2]
 CC -1- TISSUE SPECIFICITY: MAMMARY-SPECIFIC. OVER-EXPRESSED IN BREAST
 CC CANCER.
 CC -1- SUBSIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
 CC SUBFAMILY.
 CC -----

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U3147; AAC50375.1; -
 DR EMBL: AF015224; AAC39608.1; -
 DR Genew: HGNC:7050; SCGB2A2.
 DR MIM: 605562; -
 DR InterPro: IPR003627; Mamgb/prostatn.
 DR InterPro: IPR000329; Uteroglobln.
 DR Pfam: PF01099; Uteroglobln; 1.
 DR ProDom: PD029354; Mamgb/prostatn; 1.
 DR SMART: SM00096; URG; 1.
 DR PROSITE: PS00403; UTEROGLOBIN_1; FALSE_NEG.
 DR PROSITE: PS00404; UTEROGLOBIN_2; 1.
 KW Signal; Glycoprotein.
 FT SIGNAL 1 18
 FT CHAIN 19 93 POTENTIAL.
 FT CARBOHYD 53 53 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT SEQUENCE 93 AA; 10499 MW; 2896E8C43BF05312 CRC64;
 Query Match 12.6%; Score 12; DB 1 Length 93;
 Best Local Similarity 100.0%; Pred. No. 3.2e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MKLLVLMALAL 12
 Db 1 MKLLVLMALAL 12
 Oy 1 MKLLVLMALAL 12
 Db 1 MKLLVLMALAL 12
 RESULT 3
 DP02_YEAST STANDARD; PRT; 705 AA.
 ID DP02_YEAST
 AC P38121;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE DNA polymerase alpha subunit B (P86 subunit)
 GN POL12 OR YBL035C OR YBL0414.
 GN Saccharomyces cerevisiae (Baker's yeast).
 OS Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RX NCBI_TaxID=4932;
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RA MEDLINE-94378718; PubMed-8091857;
 RA Skala J., van Dyck L., Purnelle B., Goffeau A.;
 RT "The sequence of an 8.8 kb segment on the left arm of chromosome II from Saccharomycetes cerevisiae reveals four new open reading frames including homologs of animal DNA polymerase alpha-primases and bacterial GTP cyclohydrolase II.";
 RT Yeast 10:S13-S24(1994).
 RL [2]
 RP PHOSPHORYLATION.
 RX MEDLINE-96224072; PubMed-8621497;
 RA Ferrari M., Lucchini G., Plevani P., Folini H.;
 RT "Phosphorylation of the DNA polymerase alpha-primase B subunit is dependent on its association with the p180 polypeptide.";
 RT J. Biol. Chem. 271:8661-8666(1996).
 RL [2]
 CC -1- FUNCTION: MAY PLAY AN ESSENTIAL ROLE AT THE EARLY STAGE OF CHROMOSOMAL DNA REPLICATION BY COUPLING THE POLYMERASE ALPHA/PRIMASE COMPLEX TO THE CELLULAR REPLICATION MACHINERY (BY SIMILARITY).
 CC -1- SUBUNIT: DNA POLYMERASE ALPHA-PRIMASE IS A FOUR SUBUNIT ENZYME (SUBUNITS A, B, C AND D), WHICH IS ASSEMBLED THROUGHOUT THE CELL CYCLE. THE LARGEST SUBUNIT (SUBUNIT A) HAS DNA POLYMERASE ACTIVITY, THE TWO SMALLEST SUBUNITS (SUBUNITS C AND D) HAVE DNA

```

CC      PRIMASE ACTIVITY. SUBUNIT B BINDS TO SUBUNIT A (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER.
CC      -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE ALPHA SUBUNIT B FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR      EMBL: Z35796; CAA84855.1; -.
CC      DR      EMBL: X74738; CAA52761.1; -.
CC      DR      PIR: S45769; S45769
CC      SCD: S0000131; POL12.
CC      KW      DNA replication; Nuclear protein; Phosphorylation.
CC      FT      DOMAIN 80
CC      FT      POLY-SER 83
CC      SQ      SEQUENCE 705 AA: 78774 MW: 79F06D12F6979637 CRC64;
CC      -----
CC      Query Match          9.5%; Score 9; DB 1; Length 705;
CC      Best Local Similarity 100.0%; Pred. No. 0.19;
CC      Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC      -----
QY      44 ELLOEFIDS 52
QY      |||||
DB      411 ELLOEFIDS 419

RESULT 4
NRFE_ECOLI
ID      NRFE_ECOLI          STANDARD:      PRT: 552 AA.
AC      P32710.
DT      01-OCT-1993 (Rel. 27, Created)
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Cytochrome c-type biogenesis protein nrfe.
GN      NRFE OR BA074.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
OX      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94335626; Pubmed=8057835;
RX      Hussain H.A., Grove J., Griffiths L., Busby S., Cole J.;
RT      "A seven-gene operon essential for formate-dependent nitrite
RT      reduction to ammonia by enteric bacteria.";
RL      Mol. Microbiol. 12:153-163(1994).
RN      [2]
RN      SEQUENCE FROM N.A.
RP      STRAIN=K12 / MG1655.
RC      MEDLINE=94089392; Pubmed=8265357;
RA      Blatter F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA      Daniels D.L.;
RT      "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT      region from 89.2 to 92.8 minutes.";
RL      Nucleic Acids Res. 21:5408-5417(1993).
CC      -1- FUNCTION: MAY BE REQUIRED FOR THE BIOGENESIS OF C-TYPE
CC      CYTOCHROMES. POSSIBLE SUBUNIT OF A HEME LYASE.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC      (potential).
CC      -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCL1/NRFE/CCSA FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----

```

```

DR EMBL: X72298; CAA51045.1;
DR EMBL: U00006; AAC43168.1;
DR EMBL: AE000481; AAC3457.1;
DR Ecogene: EG11948; nfe.
DR InterPro: IPR002541; Cytc_asm.
DR InterPro: IPR003567; Cyt_c_b10g.
DR InterPro: IPR003568; Cyt_c_b10g_Ccmf.
DR Pfam: PF01578; Cytc_asm.1.
DR PRINTS: PR01410; CCBIOGENESIS.
DR TIGRfams: TIGR00353; nfe.1.
KW Cytochrome c-type biogenesis; Transmembrane; Inner membrane;
Complete proteome.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 77 97 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 222 242 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
FT TRANSMEM 376 396 POTENTIAL.
FT TRANSMEM 399 419. POTENTIAL.
FT TRANSMEM 525 545 POTENTIAL.
FT CONFLICT 32 32 V -> L (IN REF. 2).
FT CONFLICT 38 38 A -> G (IN REF. 2).
FT CONFLICT 190 190 S -> C (IN REF. 2).
FT CONFLICT 218 218 L -> W (IN REF. 2).
FT CONFLICT 220 226 RLVLVGS -> GWMFMDP (IN REF. 2).
FT CONFLICT 455 455 A -> G (IN REF. 2).
FT CONFLICT 481 481 A -> R (IN REF. 2).
SQ SEQUENCE 552 AA; 60884 MW; A0F3410AE3D6416B CRC64;

Query Match 8.4%; Score 8; DB 1; Length 552;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LMLALLL 14
Db 114 LMLALLL 121

RESULT 5
POLR_ASGVP STANDARD: PRT: 2105 AA.
AC P36309;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: RNA replicase (EC 2.7.7.48); Helicase; Coat protein].
OS Apple stem grooving virus (strain P-209) (ASGV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Capillivirus.
OX NCBI_TaxID=36402;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93033164; PubMed=1413530;
RA Yoshikawa N., Sasaki E., Kato M., Takahashi T.;
RT "The nucleotide sequence of apple stem grooving capillivirus genome.";
RL Virology 191:98-105(1992).
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).
CC -!- PTR: THE N-TERMINAL OF THE COAT PROTEIN IS BLOCKED.
CC -!- MISCELLANEOUS: THE COAT PROTEIN IS LOCATED IN THE CARBOXY-TERMINAL REGION OF THIS POLYPROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

```

CC EMBL: D14995; BAA03639.1; -
 DR PIR: A44059; A44059.
 DR MEROPS: C35.001; -
 DR InterPro: IPR001788; RNA_dep-RNAPol2.
 DR InterPro: IPR000606; Viral_helicase1.
 DR Pfam: PF01443; Viral_helicase1.1.
 DR RNA-directed RNA polymerase: Transferase; Polypeptide; ATP-binding;
 KM Coat protein; Helicase.
 FT NP_BIND 781 788 ATP (POTENTIAL).
 FT DOMAIN 1364 1453 POLYMERASE ACTIVE SITE.
 SO SEQUENCE 2105 AA; 241240 MW; 331291A5FA137131 CRC64.

Query Match 8.4%; Score 8; DB 1; Length 2105;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 KELLOEFT 50
 DB 1223 KELLOEFT 1230

RESULT 6

UCN2_HUMAN STANDARD: PRT: 112 AA.
 AC 096RP3; 09BUG0.
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Urocorin II precursor (Ucn II) (Stresscopin-related peptide)
 GN (Urocorin-related peptide).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RX MEDLINE=21227098; PubMed=11329063;
 RA Hsu S.Y., Hsueh A.J.W.;
 RT "Human stresscopin and stresscopin-related peptide are selective
 RT ligands for the type 2 corticotropin-releasing hormone receptor.";
 RL Net. Med. 7:605-611(2001).
 RN [12]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Skin, and uterus;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Suppress food intake, delays gastric emptying and
 CC decreases heat-induced edema. Might represent an endogenous ligand
 CC for maintaining homeostasis after stress.
 CC -!- SUBUNIT: Binds with high affinity to CRF receptors 2-alpha and 2-
 CC beta.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO THE SAUVAGINE/CORTICOTROPIN-RELEASING
 CC FACTOR/ROTEININ I FAMILY OF PEPTIDES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF320560; AAK52672.1; -
 DR EMBL: BC002647; AAH02647.1; ALT_INIT.
 DR EMBL: BC022096; AAH22096.1; -
 DR MIM: 605902; -
 DR InterPro: IPR000187; Corticotiberin.
 DR SMART: SM00039; CRF. 1.
 KM Hormone; Amidation; Signal.

FT SIGNAL 1 19 POTENTIAL.
 FT PROPEP 20 70
 FT CHAIN 72 112 UROCORIN II.
 SO SEQUENCE 112 AA; 12146 MW; 0EFE29A3463721D4 CRC64;

Query Match 7.4%; Score 7; DB 1; Length 112;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 LNWVLM 9
 DB 8 LNWVLM 14

RESULT 7

GIDB_CLOAB STANDARD: PRT: 239 AA.
 ID GIDB_CLOAB
 AC 097CW4;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Methyltransferase gldb (EC 2.1.1.-) (Glucose inhibited division
 DE protein B).
 GN GIDB OR CAC3732.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 ON NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP SRRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
 RA Tatunov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4838(2001).
 CC -!- FUNCTION: Probable S-adenosyl-L-methionine dependent
 CC methyltransferase specific for a sterol and/or lipid substrate (By
 CC similarity).
 CC -!- SIMILARITY: BELONGS TO THE GIDB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE007868; AAK81652.1; -
 DR InterPro: IPR003682; gldb.
 DR InterPro: IPR000051; SAM_bind.
 DR Pfam: PF02527; Gldb. 1.
 DR TIGRFAMs: TIGR00138; gldb. 1.
 KM TIGRFAMs: Methyltransferase; Complete proteome.
 SO SEQUENCE 239 AA; 27054 MW; 399696A4E1E4E15 CRC64;

Query Match 7.4%; Score 7; DB 1; Length 239;
 Best Local Similarity 100.0%; Pred. No. 8.6;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 YKELLOE 48
 DB 29 YKELLOE 35

RESULT 8

TYRA_HAEIN STANDARD: PRT: 377 AA.
 ID TYRA_HAEIN
 AC P43902;

```

DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE T-protein [includes: Chorismate mutase (EC 5.4.99.5) (CM): Prephenate
DE dehydrogenase (EC 1.3.1.12) (PDH)].
GN TYRA OR H11290.
OS Haemophilus influenzae.
OC Bacteria: Proteobacteria: gamma subdivision: Pasteurellaceae:
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Kleischmann A.R., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervase A.R., Bolt C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weiman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ulfendick T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fiteberg L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD Rd."
RL Science 269:496-512(1995).
CC -I- CATALYTIC ACTIVITY: Prephenate + NAD(+) = 4-hydroxyphenylpyruvate
CC -I- CATALYTIC ACTIVITY: Prephenate + NAD(+) = 4-hydroxyphenylpyruvate
CC + CO(2) + NADH.
CC -I- PATHWAY: Tyrosine biosynthesis.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32809; AAC22939.1;
DR TIGR: H11290;
DR InterPro: IPR002701; Chorismate_mut.
DR InterPro: IPR003099; PDH.
DR Pfam: PF01817; Chorismate_mut; 1.
DR Pfam: PF02153; PDH; 1.
DR Tyrosine biosynthesis; Isomerase; Oxidoreductase; NAD:
KW Multifunctional enzyme; Complete proteome.
FT DOMAIN 1 377 CHORISMATE MUTASE.
FT DOMAIN 2 377 PREPHENATE DEHYDROGENASE.
SQ SEQUENCE 377 AA; 43022 MW; 319722CFFDE5791 CRC64;

Query Match 7.4%; Score 7; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 31 KTINSDI 37
GN 111111
DB 94 KTINSDI 100

RESULT 9
PROA_SYNY3 STANDARD; PRT: 433 AA.
AC P54902;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41)
DE (Glutamate-5-semialdehyde dehydrogenase) (Glutamy] gamma-semialdehyde
DE dehydrogenase) (GSA dehydrogenase)
DE PROA OR SL0373.
GN Synechocystis sp. (strain PCC 6803).

```

```

OC Bacteria: Cyanobacteria: Chroococcales: Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sakaki T., Miyajima N.,
RA Sugitara M., Tabata S.,
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
CC -I- FUNCTION: CATALYZES THE NADPH DEPENDENT REDUCTION OF L-GAMMA-
CC GLUTAMYL 5-PHOSPHATE INTO L-GLUTAMATE 5-SEMIALDEHYDE AND
CC PHOSPHATE. THE PRODUCT SPONTANEOUSLY UNDERGOES CYCLIZATION TO FORM
CC 1-PYRROLINE-5-CARBOXYLATE.
CC -I- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
CC NAD(+) = L-gamma-glutamyl 5-phosphate + NADPH.
CC -I- PATHWAY: proline biosynthesis; second step.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -I- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D64002; BAA10399.1;
DR InterPro: IPR002086; Aldehyde_dehydr.
DR InterPro: IPR000965; Gglut_pp_reduct.
DR Pfam: PF00171; aldehyd; 1.
DR TIGR: TIGR00407; PROA; 1.
DR PROSITE: PS01223; PROA; 1.
KW Oxidoreductase; Proline biosynthesis; NADP: Complete proteome.
SQ SEQUENCE 433 AA; 46350 MW; 41C590A10DC90283 CRC64;

Query Match 7.4%; Score 7; DB 1; Length 433;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 54 AAEEAMG 60
GN 111111
DB 59 AAEEAMG 65

RESULT 10
LEUL_BUCRP STANDARD; PRT: 518 AA.
ID LEUL_BUCRP
AC P48571;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 2-isopropylmalate synthase (EC 4.1.3.12) (Alpha-isopropylmalate
DE synthase) (Alpha-IPM synthetase).
DE LEUA.
OS Buchnera aphidicola (subsp. Rhopalosiphum padi).
OG plasmid PRE.
OC Bacteria: Proteobacteria: gamma subdivision: Buchnera.
OX NCBI_TaxID=98793;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95333198; PubMed=7608990;
RA Bracho A.M., Martinez-Torres D., Noya A., Latorre A.;
RT "Discovery and molecular characterization of a plasmid localized in
RT Buchnera sp. bacterial endosymbiont of the aphid Rhopalosiphum
RT padi."
RL J. Mol. Evol. 41:67-73(1995).
CC -I- FUNCTION: Catalyzes the condensation of the acetyl group of
CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).

```

CC -1- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA =
 CC acetyl-CoA + 3-methyl-2-oxobutanate + H₂O.
 CC -1- PATHWAY: Leucine biosynthesis; first step.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
 CC SYNTHASE FAMILY. LEUA 1 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X71612; CA50615.1; -
 DR InterPro: IPR002034; AIPM/Hcit_synth.
 DR InterPro: IPR000891; HMGL-1like.
 DR Pfam: PF00682; HMGL-1like. 1.
 DR TIGRfams: TIGR00973; leua_bact; 1.
 DR PROSITE: PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
 DR PROSITE: PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
 KM Leucine biosynthesis; Lysase; Plasmid.
 SO SEQUENCE 518 AA: 57336 MW: C889C91D5720BA40 CRC64:
 Query Match 7.4%: Score 7; DB 1; Length 518;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 Oy 53 DAAAEAM 59
 Db 82 DAAAEAM 88
 RESULT 11
 ARS2_DROME STANDARD: PRT; 943 AA.
 AC 05V9K7;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Arsenite-resistance protein 2 homolog.
 GN CG7843.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RX MDLINE=20196006; Pubmed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
 RA Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burlingame D., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Haisler N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reineert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.H., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RC STRAIN-Berkeley;
 RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celisner S.E.,
 RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
 RA Harris N.L., Krommiller B., Marshall B., Millburn G.H., Richter J.,
 RA Russo S., Searle S.M.J., Smith E., Shu S., Slutskiy F.,
 RA Whitfield C.J., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: a long form (shown here) and a
 CC short form: are produced by alternative splicing.
 CC -1- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE003784; AAM68343.1; -
 DR EMBL: AE003784; AAM68343.1; -
 DR FlyBase: FBgn0033062; CG7843.
 KM Hypothetical protein; Alternative splicing.
 FT VARSPLIC 47 50 MISSING (IN SHCRT ISOFORM).
 SO SEQUENCE 943 AA: 107221 MW: OC1AF09E02E8AB0B CRC64:
 Query Match 7.4%: Score 7; DB 1; Length 943;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 Oy 29 VEKRTINS 35
 Db 297 VEKRTINS-303
 RESULT 12
 GUNE_MYCTU STANDARD: PRT; 994 AA.
 ID GUNE_MYCTU
 AC Q10379;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable glutamate-aminonit-lyase adenylyltransferase (EC 2.7.7.42)
 GN (Glutamate-synthetase adenylyltransferase) (ATAS5).
 DE GUNE OR RV2221C OR MT2279 OR MYC190.32C OR MYC427.02C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 CC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NX NCBI_TaxID=1773;
 RN [1]

1

SEQUENCE FROM N.A.

RA STRAIN-H37RV;
 RA MEDLINE-98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. II, Tekala F.,
 RA Badock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellern S., Squares R.,
 RA Sulten J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RA complete genome sequence";
 RA Nature 393:537-544(1998).
 RA [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri R., Gill J., Mikula A.,
 RA Bishai W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RA laboratory strains";
 RA Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 RL FUNCTION: ADENYLYLATION AND DEADENYLYLATION OF GLUTAMINE
 CC SYNTHETASE (BY SIMILARITY).
 CC -i- CATALYTIC ACTIVITY: ATP + [L-glutamate:ammonia ligase (ADP-
 CC forming)] -> diphosphate + adenylyl-[L-glutamate:ammonia ligase
 CC (ADP-forming)].
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: 270692; CAA94664.1;
 DR EMBL: AE007073; AAK46564.1;
 DR TIGR: MT2279;
 DR Tuberculist: RV2221c;
 DR InterPro: IPR005190; GINE.
 DR Pfam: PF03710; GINE; 2.
 DR Transferrase; Repeat; Complete proteome.
 FT CONFLICT 676 T->S (IN REF. 2).
 FT SEQUENCE 994 AA: 109137 MW: 8E5EA5D63CCE25E0 CRC64:
 SQ
 Query Match 7.4%: Score 7; DB 1; Length 994;
 Best Local Similarity 100.0%: Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VMLAAL 12
 DB 157 VMLAAL 163

RESULT 13
 Y8KD_STRLI
 ID Y8KD_STRLI STANDARD: PRT: 79 AA.
 AC P22402;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-AUG-1991 (Rel. 19, Last annotation update)
 DE Hypothetical 8.4 kDa protein (ORF 79).
 OS Streptomyces lividans.
 OC Plasmid pJ101.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1916;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-89008081; PubMed-3170481;
 RA Kendall K.J., Cohen S.N.;
 RT "Complete nucleotide sequence of the Streptomyces lividans Plasmid
 RT pJ101 and correlation of the sequence with genetic properties";
 RL J. Bacteriol. 170:4634-4651(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M21778; NOT_ANNOTATED_CDS.
 DR PIR: B30924; B30924.
 DR Hypothetical protein; Plasmid
 SQ SEQUENCE 79 AA: 8446 MW: 36DE428DA065011E C1C64:
 Query Match 6.3%: Score 6; DB 1; Length 79;
 Best Local Similarity 100.0%: Pred. No. 36;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 LAALL 14
 DB 16 LAALL 21

RESULT 14
 TFPI_RAT
 ID TFPI_RAT STANDARD: PRT: 81 AA.
 AC Q63467;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trefoil factor 1 precursor (PS2 protein).
 GN TFPI OR PS2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Stomach;
 RX MEDLINE-9643113; PubMed-8836141;
 RA Itoh H., Tomita M., Uchino H., Kobayashi T., Kataoka H., Sekiya R.,
 RA Nawa Y.;
 RA "cDNA cloning of rat ps2 peptide and expression of trefoil peptides
 RT in acetic acid-induced colitis";
 RT Biochem. J. 318:939-944(1996).
 RL -----
 RL FUNCTION: P-TYPE DOMAIN PEPTIDES MAY BE INVOLVED IN THE
 CC MAINTENANCE OF MUCOSAL INTEGRITY. THE EXACT FUNCTION OF PS2
 CC HAS NOT BEEN DETERMINED YET.
 CC -----
 CC -i- SUBCELLULAR LOCATION: Secreted.
 CC -i- SIMILARITY: CONTAINS 1 P-TYPE (TREFOIL) DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D83231; BAAL1857.1;
 DR HSP: P04155; IP52.
 DR InterPro: IPR000519; P-trefoil.
 DR Pfam: PF00088; trefoil; 1.
 DR PRINTS: PR00680; PTREFOIL.
 DR SMART: SM00018; P; 1.
 DR PROSITE: PS00025; P-TREFOIL; 1.
 KM Growth factor; Signal.
 FT SIGNAL 1 21 POTENTIAL.

```

FT CHAIN 22 81 TREFOIL FACTOR 1.
FT MOD_RES 22 22 PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT DOMAIN 27 68 P-TYPE.
FT DISULFID 28 54 BY SIMILARITY.
FT DISULFID 38 53 BY SIMILARITY.
FT DISULFID 48 65 BY SIMILARITY.
FT DOMAIN 76 78 GLU-RICH (ACIDIC).
SQ SEQUENCE 81 AA: 9155 MW: 56740550406ECF25 CRC64:

Query Match 6.3% Score 6: DB 1: Length 81:
Best Local Similarity 100.0% Pred. No. 37:
Matches 6: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Oy 5 MWLMLA 10
    |||||
Db 11-MWLMLA 16

RESULT 15
PSC3_RAT
ID PSC3_RAT STANDARD: PRT: 95 AA.
AC P02780: 063463:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Prostatic steroid-binding protein C3 chain precursor (Prostatein
DE peptide C3).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83082848: PubMed=6294095:
RA Parker M.G., White R., Hurst H., Needham M., Tilly R.;
RT "Prostatic steroid-binding protein. Isolation and characterization of
RT C3 genes.";
RL J. Biol. Chem. 258:12-15(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=83238526: PubMed=6190812:
RA Viskochil D.H., Perry S.T., Lea O.A., Stafford D.W., Wilson E.M.,
RA French F.S.;
RT "Isolation of two genomic sequences encoding the Mr = 14,000 subunit
RT of rat prostatein.";
RL J. Biol. Chem. 258:8861-8866(1983).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92165796: PubMed=1537831:
RA Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M.,
RA French F.S.;
RT "Response elements of the androgen-regulated C3 gene.";
RL J. Biol. Chem. 267:4456-4466(1992).
RN [4]
RP ERRATUM.
RX MEDLINE=92218467: PubMed=1339454:
RA Tan J.A., Marschke K.B., Ho K.C., Perry S.T., Wilson E.M.,
RA French F.S.;
RL J. Biol. Chem. 267:7958-7958(1992).
RN [5]
RP SEQUENCE OF 19-95.
RX MEDLINE=81188769: PubMed=7014218:
RA Peeters B., Rombauts W., Mous J., Heyns W.;
RT "Structural studies on rat prostatic binding protein. The primary
RT structure of its glycosylated component C3.";
RL Eur. J. Biochem. 115:115-121(1981).
CC -1- FUNCTION: PART OF PROSTATEIN WHICH IS THE MAJOR SECRETORY
CC GLYCOPROTEIN OF VENTRAL PROSTATE GLAND. STEROID-BINDING PROTEIN;
CC CAN BIND NON-POLAR STEROIDS, CHOLESTEROL AND A GROUP OF SMALL
CC PROLINE-RICH PEPTIDES.
CC -1- SUBUNIT: PROSTATEIN IS COMPOSED OF THREE DIFFERENT PEPTIDES CALLED
CC C1, C2 AND C3. THESE FORM COVALENT C1:C3 (F) AND C2:C3 (S)
CC HETERODIMERS WHOSE NONCOVALENT ASSOCIATION FORMS TETRAMERIC

```

```

CC (C1:C3/C3:C2) PROSTATEIN MOLECULES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: VENTRAL PROSTATE.
CC -1- INDUCTION: ANDROGEN DEPENDENT. AS SHOWN BY THE DECREASE IN THE
CC LEVEL OF THE PROTEIN FOLLOWING CASTRATION.
CC -1- MISCELLANEOUS: C3 IS ENCODED BY TWO DIFFERENT GENES.
CC -1- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01263: CAA24577.1; -;
DR EMBL: M71245: AAA41965.1; -;
DR PIR: A03250: BORT3.
DR InterPro: IPR003627: Mamgb/prostatn.
DR InterPro: IPR000329: Uteroglobln..
DR Pfam: PF01099: Uteroglobln: 1.
DR ProDom: PD029354: Mamgb/prostatn: 1.
DR PROSITE: PS00403: UTEROGLOBIN_1: FALSE_NEG.
DR PROSITE: PS00404: UTEROGLOBIN_2: 1.
KW Signal: Glycoprotein; Steroid-binding.
FT SIGNAL 1 18
FT CHAIN 19 95 PROSTATIC STEROID-BINDING PROTEIN C3
FT CHAIN.
FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .).
FT CONFLICT 53 53 D -> A (IN REF. 3).
FT CONFLICT 79 79 G -> S (IN REF. 2).
SQ SEQUENCE 95 AA: 10730 MW: F7F7FLAOC882E375 CRC64:

Query Match 6.3% Score 6: DB 1: Length 95:
Best Local Similarity 100.0% Pred. No. 42:
Matches 6: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Oy 62 FKOCFL 67
    |||||
Db 63 FKOCFL 68

```

Search completed: January 25, 2003, 13:40:32
 Job time : 12 secs